

# SCIENTIFIC REPORTS



OPEN

## Signatures of balancing selection in toll-like receptor (TLRs) genes – novel insights from a free-living rodent

Agnieszka Kloch<sup>1</sup>, Marius A. Wenzel<sup>2</sup>, Dominik R. Laetsch<sup>3,4</sup>, Olek Michalski<sup>5</sup>, Renata Welc-Falęciak<sup>6</sup> & Stuart B. Piertney<sup>2</sup>

Selective pressure from pathogens is considered a key selective force driving the evolution of components of the immune system. Since single components of the immune system may interact with many pathogens, and single pathogens may be recognized by multiple components of the immune system, gaining a better understanding of the mechanisms of parasite-driven selection requires the study of multiple genes and pathogens. Toll-like receptors (TLRs) are a large gene family that code for antigen-presenting components of the innate immune response. In the present paper we characterize polymorphism and signatures of selection in seven TLRs in free-living bank voles *Myodes glareolus*. We report the first evidence of balancing selection in several TLR genes, supported by positive values of Fu and Li's D\* in TLR2 and TLR5, and positive values of Tajima's D in LRR regions within TLR1 and TLR2. We further found significant associations between amino-acid alleles of TLR1 and TLR5 and susceptibility to infection with the blood pathogen *Bartonella*. Interestingly, selection patterns in TLRs presenting virus-derived motifs (TLR7 and TLR9) differed considerably from those interacting with bacterial PAMPs. In contrast to the highly variable TLRs presenting bacterial motifs, TLR7 and TLR9 had low polymorphism and displayed signatures of directional selection. These findings suggest different functional responses across the TLR gene family and highlight the complexity of parasite-driven selection.

Host-pathogen relationships have long been in the focus of evolutionary biology as a prominent example of an evolutionary arms race. In the wild, animals are exposed to a spectrum of pathogens, and each pathogen employs different mechanisms to escape from host defence. As a result, components of the immune system are expected to evolve under parasite-driven selection. The strength and direction of this selection depends on the pathogen repertoire, as a single component of the immune system may interact with many pathogens, while single pathogens may be recognized by multiple components of the immune system. Given the complexity of these interactions, understanding the mechanisms of pathogen resistance in the wild requires the study of multiple loci and multiple pathogens<sup>1</sup>.

However, much of what we know about mechanisms of parasite-driven selection has been based on a single locus, namely major histocompatibility complex (MHC). MHC constitutes part of the adaptive immune response, which provides specific but delayed reaction to an infection. Moreover, most variance among hosts in their resistance against pathogens may be explained by factors other than MHC repertoire<sup>2</sup>, and other immunity genes have been shown to play a crucial role in response against helminth infections in free-living vertebrates (e.g.<sup>3–5</sup>).

Toll like receptors (TLRs) are the key component of the innate immunity. They recognize pathogen-associated molecular patterns (PAMP) derived from microorganisms such as viruses, bacteria, protozoa, and fungi. After binding to a motif, TLR molecule triggers inflammation and activates signalling pathways which initiate later

<sup>1</sup>University of Warsaw, Department of Ecology, Faculty of Biology, Biological and Chemical Research Centre, ul. Zwirki i Wigury 101, 02-089, Warszawa, Poland. <sup>2</sup>University of Aberdeen, Institute of Biological and Environmental Sciences, Tillydrone Avenue, Aberdeen, AB24 2TZ, UK. <sup>3</sup>The James Hutton Institute, Dundee, DD2 5DA, United Kingdom. <sup>4</sup>University of Edinburgh, Institute of Evolutionary Biology, Edinburgh, EH9 3JT, UK. <sup>5</sup>Data Analysis & Visualisation, ul. Przy Agorze 5a m. 98, 01-960, Warszawa, Poland. <sup>6</sup>University of Warsaw, Department of Parasitology, Faculty of Biology, ul. Miecznikowa 1, 02-096, Warszawa, Poland. Correspondence and requests for materials should be addressed to A.K. (email: [akloch@biol.uw.edu.pl](mailto:akloch@biol.uw.edu.pl))

stages of the immune response, including the adaptive response. Clinical studies have shown that polymorphism within TLR genes, particularly in the part of the molecule interacting with PAMP, affects susceptibility to diseases in humans – for instance, a single nucleotide polymorphism (SNP) in the TLR2 gene has been attributed to susceptibility to tuberculosis, cytomegalovirus, and borreliosis, whereas SNPs associated with susceptibility to hepatitis C virus have been identified in TLR7 (see the work reviewed in<sup>6</sup>). Associations between TLR variants and susceptibility have also recently been found in TLR2 and TLR4<sup>4,7</sup> in free living species.

Selection acting on TLR genes has seldom been studied in a microevolutionary context, and numerous studies that have focused on macro-evolutionary processes have provided no clear picture. Positive selection has been reported in primates<sup>8,9</sup>, and across several mammalian taxa<sup>10</sup>. In birds, TLR evolution was found to be dominated by stabilizing selection and slow rates of non-synonymous substitutions<sup>11</sup>. In two closely related newt species, purifying selection was observed over a long time scale but evidence for positive selection was found on a medium timescale<sup>12</sup>. Across several rodent species, signatures of positive selection were detected in PAMPs of TLR2 but the remaining part of the protein seemed to evolve under purifying selection<sup>13</sup>. TLRs presenting bacterial ligands tend to differ in their evolutionary pattern from TLRs interacting with viral motifs<sup>8,14</sup>, yet this finding was not supported in all taxa studied<sup>10</sup>.

Although immunity genes affected by parasite-driven selection are likely to evolve under balancing selection, to date this has only been reported in TLR1, 6, and 10 in humans, in the European population<sup>15</sup>. It was not observed in other populations, which authors have explained in terms of different pathogen pressures, yet the effect of pathogens was not directly included in the analysis. The effect of balancing selection on genomic regions is often subtle compared to positive selection<sup>16</sup>, and may be difficult to detect over a long evolutionary scale<sup>17</sup>. To better understand micro-evolutionary processes affecting polymorphism within TLR genes, studies on a small spatial or temporal scale are necessary. Hence, in the current paper we tested signatures of selection as well as associations between infections and genetic variants. The sites studied differ in terms of pathogen communities, and these differences are stable over time<sup>18</sup>. In our previous work in this study system, we found local, contrasted associations between MHC DRB beta chain variants and susceptibility to infection with the nematode *Aspiculuris tetraptera*<sup>19</sup>.

## Methods

**Material and pathogen screening.** DNA samples were obtained from 90 bank voles (*Myodes glareolus*) collected at two sites (Urwitałt and Pilchy, NE Poland) in a single trapping session in 2005. These individuals were previously genotyped in MHC-DRB and 6 microsatellite loci<sup>19</sup>. The helminth infection level was determined based on autopsy material; for details see<sup>19</sup>. In the present study we also considered infection with microparasites. The intestinal protist *Cryptosporidium sp.* was identified in faecal smears using the Ziehl-Nielsen staining technique<sup>20</sup>. The blood pathogens *Bartonella sp.* and *Babesia sp.* were identified by PCR using primers BhCS.781p and BhCS.1137n for *Bartonella sp.*<sup>21</sup>, and nested reaction with primers CRYPTO R and CRYPTO F<sup>22</sup>, and Bab1 and Bab4 for *Babesia*<sup>23</sup>. Sequences of the primers and reaction conditions are given in Table S1. The remaining blood pathogens – the bacteria *Haemobartonella sp.* and the protists *Trypanosoma sp.* and *Hepatozoon sp.* – were microscopically identified based on size, shape and colouration upon Giemsa-staining. The list of the identified species, along with their prevalence, is given in Table S2.

All procedures were performed in compliance with the Animal Protection Act of 1997 (Dz.U. 1997 no. 111 pos. 724), followed the guidances of the National Ethics Committee for Experimentation on Animals (Poland), and were approved by the Local Ethical Committee no. 1 in Warsaw, decision no. 280/2003.

**Library preparation.** Primers amplifying toll-like receptors TLR1, TLR4, TLR5, TLR6, TLR7 and TLR9 were designed in conserved regions of aligned rodent mRNA sequences obtained from the NCBI GenBank. Despite several attempts, we were not able to design primers amplifying TLR3 and TLR8. TLR2 was amplified using two pairs of overlapping primers described in<sup>13</sup>.

To minimize the PCR error rate, we used high fidelity Phusion polymerase (Thermo Scientific). The PCR mix contained: 5x polymerase buffer, 200 μM of each dNTP, 0.5 μM of each primer, 0.3 U of polymerase, and 10–50 ng of genomic DNA. Details of both primers' melting temperature and annealing time are given in Table S3.

Amplified fragments were pooled for each individual in similar quantities based on intensities of resolved bands by agarose gel electrophoresis<sup>24</sup>, and purified twice using CleanUp kit (Aabiot). The library was prepared using Nextera XT DNA according to the manufacturer's protocol and sequenced in a single run on Illumina MiSeq using MiSeq Reagent Kit v3 (150 cycles) producing 2 × 75 bp reads. The quality of the read datasets was assessed using the FastQC tool (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

**Assembly and genotyping.** To construct a reference for read mapping, an unannotated bank vole genome assembly was retrieved from the NCBI GenBank (PRJNA290429). Mouse transcripts (accession numbers NM\_001276445, NM\_011905, NM\_021297, NM\_016928, NM\_011604, NM\_001290755, NM\_031178) of the studied TLR genes were mapped against the bank vole genome using the -f same option of gmap tool (<http://research-pub.gene.com/gmap>)<sup>25</sup>. The regions against which mouse transcripts mapped were extracted with ~1000 bp offset using getfasta function in Bedtools ver. 2.25 (<http://bedtools.readthedocs.io/en/latest>)<sup>26</sup> and used as reference in the downstream analysis.

The TLR9 locus in the bank vole genome contained two unassembled parts (-NNN-). Thus, after initial mapping as described below, we designed internal primers (TLR9\_intF 5'TCTGCCCAACCTCCATACTC3' and TLR9\_intR 5' TCTTATGGTCAGGGGTGCTC3'), which, together with the TLR9 F/R primers listed in Table S3, were used to amplify regions overlapping the -NNN- parts. The amplicons were Sanger-sequenced providing reference sequences with filled -NNN- parts that were used to construct a gap-free reference sequence for this locus.

Paired-end reads were mapped against the reference using *bwa-mem* ver. 0.5.9 (<http://bio-bwa.sourceforge.net>) with default parameters<sup>27</sup>. The resulting bam files were merged, and deduplicated using the Picard MarkDuplicates tool (<http://broadinstitute.github.io/picard>). Duplicated read-pairs constituted 87% of the reads. For the further analysis, we retained only properly mapped and paired read pairs (both reads mapped, correct insert size) with a mapping quality  $\geq 20$ , which resulted in 32 213 487 reads. The high coverage obtained ( $\sim 1000\times$ ) assured high confidence of the genotyping results.

Variants were called in two-step procedure using the FreeBayes v0.9.10-3 (<https://github.com/ekg/freebayes>)<sup>28</sup>. First, we ran FreeBayes with the following parameters: minimal fraction of alternate allele of 20% ( $-F$  0.2) as recommended<sup>29</sup>, minimum number of reads supporting alternate allele  $> 2$  ( $-\text{min-alternate-count}$  2), and minimal read coverage  $> 5$  ( $-\text{min-coverage}$  5). The results were filtered using *vcfilter* v.41 (<https://github.com/vcflib/vcflib>) specifying:  $\text{QUAL/AO} > 10$  &  $\text{DP} > 10$  &  $\text{SAF} > 0$  &  $\text{SAR} > 0$  &  $\text{RPR} > 0$  &  $\text{RPL} > 0$ . In a second step, filtered high confidence variants were used to construct haplotypes using *haplotypes-basis-alleles* function and specifying  $-\text{max-complex-gap}$  37 (i.e. half of the read length). SNPs which were not phased by Freebayes were computationally phased using PHASE (<http://stephenslab.uchicago.edu/software.html#phase>)<sup>30</sup> and resolved into haplotypes using a custom Python script (included in the Supplementary Materials).

The reading frame in each locus was identified based on alignment with mouse transcripts, which allowed each SNP to be classified as synonymous or non-synonymous. Additionally, in each locus we identified leucine-rich repeats (LRRs) using the LRRfinder tool (<http://www.lrrfinder.com>)<sup>31</sup>.

**Sequence variation and signatures of selection.** Basic measures of sequence diversity – the number of variable sites (*S*), haplotype diversity (*H<sub>d</sub>*), average number of nucleotide differences (*k*), and nucleotide diversity per site (*pi*) – were summarized in DnaSP v.5 (<http://www.ub.edu/dnasp>)<sup>32</sup>. Observed and expected heterozygosity and *F<sub>is</sub>* (inbreeding coefficient) were calculated using the R package *diveRsity*<sup>33</sup>, and the *F<sub>st</sub>* statistic was computed in Arlequin ver 3.1 (<http://cmpg.unibe.ch/software/arlequin35>)<sup>34</sup>.

In the present paper we focused on signatures of selection acting on TLR genes on a small evolutionary scale. First, we tested for deviations from Hardy-Weinberg equilibrium (HWE) using the Markov chain method in Genepop v.4 (<http://kimura.univ-montp2.fr/~rousset/Genepop.htm>)<sup>35</sup>. In DnaSP, we computed two neutrality tests based on a frequency spectrum of alleles and segregating sites: Tajima's *D*<sup>36</sup> and Fu and Li's *D*<sup>37</sup>. Tajima's *D* was also calculated using a 60 bp sliding window and 20 bp step size. The size of the window corresponded to the average size of a LRR. Using mouse sequence (*Mus musculus*) as an outgroup we computed the McDonald-Kreitman (MK) test<sup>38</sup>, which compares the rate of synonymous and non-synonymous substitutions within and between species, and allows for the detection of selection acting on a longer time scale. Unlike Tajima's *D* and Fu and Li's *D*<sup>\*</sup>, this test is relatively insensitive to demographic assumptions. Additionally, in Arlequin we computed the Ewens-Watterson test, which allows contemporary selection to be distinguished from past selection<sup>39,40</sup>.

**Associations between genetic variation and infection.** We used generalized linear models to identify associations between genetic polymorphisms and susceptibility to infections. This approach is an alternative to selection tests, and detection of significant associations between genetic variants and presence of pathogens indicates contemporary parasite-driven selection operating at the studied loci<sup>41</sup>.

First, we tested the significance of three non-genetic variables: host sex, body mass, and sampling site, which were previously shown to influence nematode load in the studied populations<sup>19</sup>. For the most frequent eight pathogen species, whose overall prevalence exceeded 20% (Table S2), we fitted models with binary infection status as the dependent variable, using GLM with binomial error distribution and logit link function in R<sup>42</sup>. Non-genetic terms that were associated with the risk of infection with  $p < 0.05$  (Table S4) were included as explanatory variables in the models for the respective pathogen species.

In a second step, we selected genetic variants to be included in the models. As recommended in<sup>43</sup>, we focused on haplotype polymorphism rather than single nucleotide polymorphism (SNP). By using haplotypes, which are combinations of all SNPs within a locus, we were able to reduce the number of explanatory variables fitted to a model without losing information about the genetic variability within the studied genes. To better account for functional rather than structural differences between variants, we translated nucleotide sequences into amino-acid sequences. Such an approach has been used in similar studies examining the functional effects of the components of the immune system on susceptibility to infections<sup>7,44</sup>. From here on in this paper, unless stated otherwise, we use the term "haplotype" to denote a unique amino-acid haplotype. For clarity, the amino acid haplotypes are denoted TLR\*aa. After translation, only two unique amino acid sequences were found both in TLR7 and TLR9. The less frequent allele of TLR7 was present only in four individuals, and of TLR9 only in one animal, thus we were not able to test for an effect of any of these loci using GLM models.

To control for relatedness between individuals and possible population structure, we computed a pairwise relationship matrix based on six previously genotyped microsatellite loci<sup>19</sup> using the *irelr* package for R (<https://github.com/andersgs/irelr>)<sup>45</sup>. We summarized the variation using principal component analysis (PCA). The first and second principal components (PC1 and PC2) were fitted to each model.

For each locus (TLR1, 2, 4, 6) and each pathogen, we constructed separate models with infection status as dependent variable. In each model the explanatory variables were as follows: (1) optional non-genetic terms as described above, (2) presence/absence of the most frequent haplotypes, (3) PC1 and PC2 from the relatedness matrix summarizing population structure. To avoid overfitting, we only considered haplotypes that occurred in at least 8 individuals ( $\sim 10\%$  of the studied animals) with the exception of models including *H. mixtum* and *H. glareoli*. Each of these nematodes was observed in only one of the studied sites: *H. mixtum* at the Urwitaht site, *H. glareoli* in Pilchy. This pattern is persistent over time, as reported in a longitudinal study of the rodent parasite fauna at these sites<sup>18</sup>. Thus, models including either *H. mixtum* or *H. glareoli* were fitted using observations from

| locus | length (bp) | n  | %    | S               | S/1000bp | h  | AA | Hd    | k     | $\pi \times 10^3$ |
|-------|-------------|----|------|-----------------|----------|----|----|-------|-------|-------------------|
| TLR1  | 1126        | 81 | 47.2 | 24              | 21.31    | 16 | 12 | 0.851 | 6.387 | 5.67              |
| TLR2  | 2351        | 85 | 99.9 | 57              | 24.25    | 32 | 26 | 0.804 | 9.116 | 3.80              |
| TLR4  | 2162        | 81 | 86.3 | 21 <sup>1</sup> | 9.71     | 16 | 11 | 0.771 | 2.708 | 1.25              |
| TLR5  | 2260        | 86 | 86.3 | 39 <sup>1</sup> | 17.3     | 19 | 10 | 0.837 | 7.403 | 3.28              |
| TLR6  | 1234        | 85 | 51.0 | 20              | 16.21    | 12 | 8  | 0.811 | 5.464 | 4.43              |
| TLR7  | 2948        | 86 | 93.6 | 5               | 1.7      | 6  | 2  | 0.347 | 0.375 | 0.13              |
| TLR9  | 2962        | 83 | 95.8 | 11              | 3.71     | 9  | 2  | 0.753 | 3.372 | 1.14              |

**Table 1.** Summary of the diversity of the studied loci. Length of the genotyped sequence in bp and % – percent of protein-coding sequence covered by genotyping are given. n – number of genotyped individuals, S – number of variable sites, h – number of haplotypes, AA – number of amino-acid alleles (unique translated sequences), Hd – haplotype diversity, k – average number of nucleotide differences,  $\pi$  – nucleotide diversity per site. <sup>1</sup>Including an indel.

the respective sites, and to these models we fitted the 5 most frequent TLR variants from each site. To control for multiple comparisons, we adjusted the p-values of genetic terms using the false discovery rate (FDR) procedure<sup>46</sup>. Following<sup>47</sup>, we considered an adjusted p-value to be highly significant at  $p < 0.05$ , and significant at  $p < 0.1$ .

To provide a quantitative measure of the strength of the tested effects, we calculated  $R^2$ , an effect size statistic in the rsq package for R<sup>48</sup>.  $R^2$  is a coefficient of determination that measures the proportion of the variance in the dependent variable which can be predicted from the independent variable.

Additionally, to complement the haplotype-oriented analysis, we tested the effects of SNPs on the risk of infection. Synonymous SNPs were filtered using the PLINK 1.9 tool (<https://www.cog-genomics.org/plink2>)<sup>49</sup>. We retained individuals with at least 30% of SNPs called, and SNPs that were in the Hardy-Weinberg equilibrium (P-value cutoff 0.01) with less than 10% missing data and with minor allele frequency greater than 0.01. Linkage disequilibrium among all SNP pairs was examined using squared allele-counts correlation ( $r^2$ ), and SNPs were thinned out such that  $r^2$  among SNP pairs within a sliding window of 50 SNPs (step-size 10 SNPs) was below 0.7. Filtered SNPs were fitted as explanatory variables to GLM models constructed in a similar manner as in the case of haplotype-based models, with pathogen presence/absence as dependent variable, and genotype at a given SNP, two PCs summarizing the relatedness among individuals, and optional non-genetic terms as explanatory variables. The effect of each SNP was tested separately for each dependent variable, with p-levels adjusted using FDR.

**Data Accessibility.** Raw reads have been stored in SRA archive, biosample no. 7414489. Haplotypes are stored in GenBank (Access nos. MF471907 – MF472008). Should the manuscript be accepted, the data on infection status and genotypes of individuals will be archived in Dryad and the data DOI will be included at the end of the article.

## Results

**Genetic polymorphism and test of selection.** Overall, we found 179 variable sites in 15043 bp across seven TLR genes, including 177 SNPs and two indels (Table 1). A three-nucleotide deletion was detected in TLR4 in position 1830 (haplotype 8) in four heterozygous individuals. In TLR5, a two-nucleotide deletion present in 17 individuals was identified in position 247 of haplotypes 2 and 16, which resulted in a frame shift mutation and a premature stop codon in position 379.

Levels of polymorphism differed considerably between the loci studied. The most polymorphic loci were TLR1 and TLR2 with over 20 variable sites per 1000 bp; in these two loci we observed the highest average number of nucleotide differences (k), and the highest nucleotide diversity per site ( $\pi$ ) (Table 1). The number of segregating sites and the nucleotide diversity per site were much lower in TLR7 and in TLR9 than in other loci, and TLR7 had considerably lower haplotype diversity compared to other loci.

We found significant deviations from the Hardy-Weinberg equilibrium in TLR4, TLR7, and TLR9 (Table 2). The  $F_{st}$  coefficient showed no consistent pattern over loci: in TLR1, 2, 4 and 6 it was lower than 0.1, suggesting a low level of population differentiation, in contrast to TLR9, where  $F_{st}$  exceeded 0.25, indicating strong differentiation. By comparison,  $F_{st}$  was 0.072 in microsatellites, and 0.069 in MHC-DRB<sup>19</sup>.

Tajima's D tests were not significant for any locus, but we found significantly positive values of Fu and Li's D\* in TLR2 and TLR5 (Table 3). These two loci were also significant at  $p < 0.02$  when Fu and Li's test was computed with a mouse sequence as an outgroup (Table S5). The McDonald-Kreitman (MK) test revealed an excess of synonymous substitutions in TLR9 (Table 3). The Ewens-Watterson results were not significant at any locus, suggesting that the effect of long-term selection did not affect the observed homozygosity.

The sliding-window Tajima's test revealed regions of positive D ( $p < 0.05$ ) in both TLR1 and TLR2, which indicates an excess of intermediate frequency alleles and may suggest balancing selection. In both loci, the regions with  $D > 0$  were located within LRR (Fig. 1). LRRs form horseshoe structures in the extracellular domain of the TLR molecules, which are crucial for the TLR function, as they form sites involved in binding with pathogen-derived motifs or with other components of the immune system interacting with TLRs. Sites with positive D were also found within LRRs in both TLR5 and TLR6 (Fig. S1) but at  $0.05 < p < 0.01$ . The D statistic was not significant in all sites with  $D < 0$ .



| locus | $F_{st}$              | $F_{is}$ | HWE        |          |        |
|-------|-----------------------|----------|------------|----------|--------|
|       |                       |          | both sites | Urwitalt | Pilchy |
| TLR1  | 0.0998 <sup>***</sup> | 0.0428   | 0.099      | 0.132    | 0.154  |
| TLR2  | 0.0569 <sup>***</sup> | -0.0091  | 0.784      | 0.497    | 0.869  |
| TLR4  | 0.0166 <sup>*</sup>   | -0.0645  | 0.017      | 0.057    | 0.043  |
| TLR5  | 0.1174 <sup>***</sup> | 0.0197   | 0.958      | 0.729    | 0.992  |
| TLR6  | 0.0662 <sup>***</sup> | 0.0422   | 0.410      | 0.258    | 0.564  |
| TLR7  | 0.1069 <sup>***</sup> | 0.4315   | <0.001     | 0.156    | 0.000  |
| TLR9  | 0.3833 <sup>***</sup> | 0.0889   | 0.027      | 0.057    | 0.074  |

**Table 2.** Genetic differences between studied sites.  $F_{st}$  – fixation coefficient (p-value given in a upper index, \*\*\* $p < 0.001$ , \*\* $p < 0.1$ , \* $p < 0.05$ ),  $F_{is}$  – inbreeding coefficient, HWE – p-value of Hardy-Weinberg equilibrium test estimated by the Markov chain method.

|      | Tajima's D           | Fu & Li D*                       | McDonald-Kreitman (NI)        | Ewens-Watterson |
|------|----------------------|----------------------------------|-------------------------------|-----------------|
| TLR1 | 1.424 <sup>ns</sup>  | 0.115 <sup>ns</sup>              | 0.768 <sup>ns</sup>           | 0.337           |
| TLR2 | -0.267 <sup>ns</sup> | <b>1.089</b> <sup>&lt;0.02</sup> | 1.469 <sup>ns</sup>           | 0.996           |
| TLR4 | -0.744 <sup>ns</sup> | -0.128 <sup>ns</sup>             | 0.641 <sup>ns</sup>           | 0.543           |
| TLR5 | 0.254 <sup>ns</sup>  | <b>1.841</b> <sup>&lt;0.02</sup> | 1.070 <sup>ns</sup>           | 0.633           |
| TLR6 | 1.331 <sup>ns</sup>  | 1.297 <sup>ns</sup>              | 0.609 <sup>ns</sup>           | 0.260           |
| TLR7 | -1.118 <sup>ns</sup> | -0.136 <sup>ns</sup>             | 2.193 <sup>ns</sup>           | 0.845           |
| TLR9 | 1.813 <sup>ns</sup>  | 0.678 <sup>ns</sup>              | <b>0.138</b> <sup>0.031</sup> | 0.290           |

**Table 3.** Results of neutrality tests. Coefficients given for each test, with the p-values in superscript. For EW test, only p-values are presented. Significant variables are marked in bold.

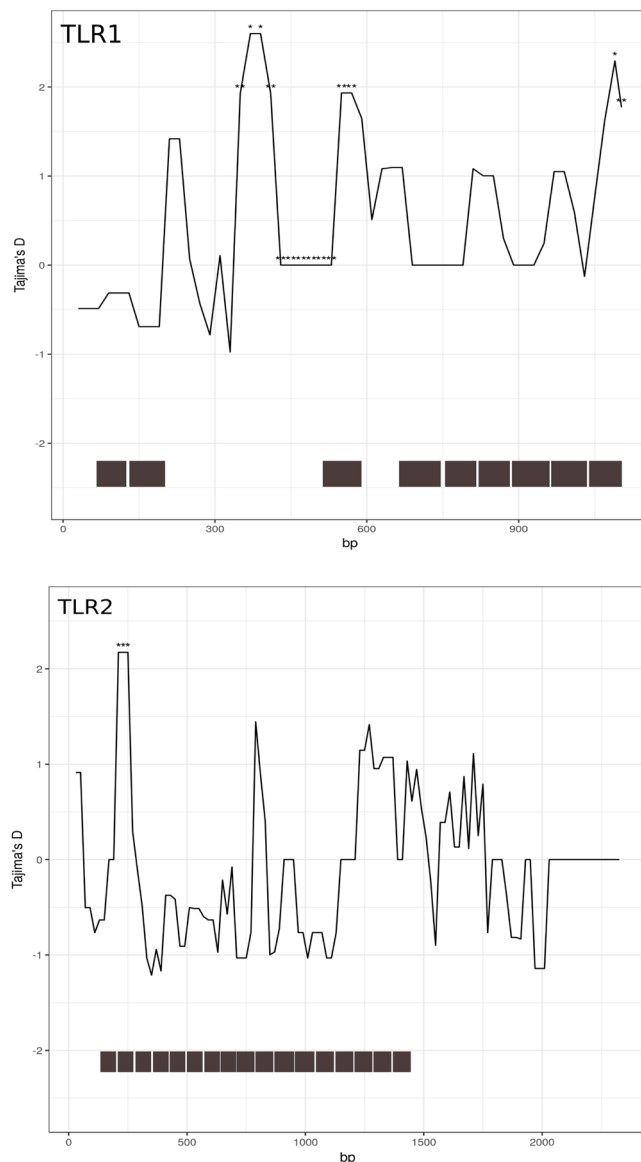
**TLR variability and susceptibility to infection.** Linear modelling revealed several associations between risk of infection and presence of the most frequent amino-acid alleles (Table S6), yet after controlling for multiple comparisons, only the effects of the TLR1\*aa10 and TLR5\*aa02 on infection with the blood pathogen *Bartonella* remained significant or marginally significant (Table 4). Animals with the allele TLR1\*aa10 were three times more often infected with *Bartonella* than those without it (adjusted  $p = 0.023$ , Table 4). Interestingly, animals with TLR1\*aa10 were five times less often infected with the nematode *A. tetraptera* than those without the allele (5.8% vs. 27.7%,  $p = 0.019$ ), yet after FDR correction this effect appeared not to be significant (adjusted  $p = 0.2$ , Table S6). The presence of the TLR5\*aa02 was associated with increased susceptibility to *Bartonella* (adjusted  $p = 0.023$ ): 41.2% animals with the allele were infected, compared to 26.8% infection among voles without it. None of the SNPs had a significant effect on the risk of infection after controlling for multiple comparisons (Table S7).

## Discussion

We have examined patterns of parasite-driven selection in the TLR gene family using a free-living rodent as a study system. By combining neutrality tests with statistical models we provide the first evidence for balancing selection acting on TLR1 and TLR5, which present bacterial PAMPs, whereas in TLR7 and TLR9, which recognize virus-derived motifs, we found weak evidence of negative and positive directional selection, respectively.

**Selection acting on TLRs presenting bacterial PAMPs.** We found no uniform pattern of selection across TLRs presenting bacterial PAMPs (TLR1, 2, 4, 5, 6). TLR4 showed significant deviation from the Hardy-Weinberg equilibrium, which resulted from an excess of heterozygotes at the site Pilchy, yet in this locus we found no other evidence for selection. In the remaining TLRs, except for TLR6, we found significant deviations from neutrality, as indicated by Fu and Li's D\* test for TLR2 and TLR5, and increased Tajima's D in parts of TLR1 and TLR2. Although, overall, Tajima's D was not significant, high significant values in parts of the molecules putatively interacting with PAMPs suggest parasite-driven selection. Positive D or D\* values are interpreted as a signature of balancing selection but they may also arise from demographic processes such as changes in population size. However, if demographic processes had affected the frequency of alleles in the studied populations, this should be similar in all loci, yet we found no consistent pattern over the genes studied<sup>50</sup>. Similarly, no uniform pattern arose from  $F_{st}$  analysis. In TLR1 and TLR5,  $F_{st}$  was close to 0.1, indicating moderate population structure, while in TLR2, 4 and 6 it was lower or similar to the value obtained from neutral markers, which was 0.072<sup>19</sup>. In Sweden, a somewhat similar discrepancy between population structure in TLR2 and in neutral loci observed in bank voles was interpreted as a evidence for local parasite-driven selection acting on this locus<sup>51</sup>. Thus, the observed differences between our sites in Poland in the TLR allele frequencies are more likely to result from selection than from demographic processes.

Macro evolutionary studies of TLR evolution have underlined the role of directional selection, either positive<sup>8,10</sup>, purifying<sup>11,13</sup> or both<sup>12</sup>. However, within a species the results have often been inconsistent: for instance in



**Figure 1.** Sliding window Tajima's D for TLR1 and TLR1. Stars indicate regions where D was significant at  $p < 0.05$ . Grey bars below the graph represent location of LRRs.

| genetic term | pathogen          | R2    | $\chi^2$ | p     | FDR   | % infected |         |
|--------------|-------------------|-------|----------|-------|-------|------------|---------|
|              |                   |       |          |       |       | with       | without |
| TLR1*aa10    | <i>Bartonella</i> | 0.144 | 10.6169  | 0.001 | 0.024 | 11.7       | 33.3    |
| TLR5*aa02    | <i>Bartonella</i> | 0.094 | 7.4701   | 0.006 | 0.066 | 41.2       | 26.8    |

**Table 4.** Genetic terms significantly associated with a risk of infection. R2 - coefficient of determination, p - p-value as fitted in a model,  $\chi^2$  - test statistics, p - p-value, FDR - adjusted p-value using false discovery rate, % infected with and without the significant genetic variant.

humans, the European population showed signatures of balancing selection<sup>15</sup>, whereas the Indian population had notably higher number of non-synonymous substitutions compared to other studied areas<sup>52,53</sup>.

Except for the human study cited above, our work is the first to demonstrate balancing selection acting on TLR genes in mammals. Parasite-driven balancing selection, either based on frequency dependent selection or rare allele advantage, has been well described in the case of MHC genes (see the work reviewed in<sup>39</sup>), although its importance for innate immunity genes has yet to be explained. To date, evidence for balancing selection has been found in cytokines IL1b, IL 2 and TNF  $\alpha$  in field voles, as indicated by significant deviation from neutrality in one (IL1b, IL 2) or several tests (TNF  $\alpha$ )<sup>43,50</sup>. In the current study, we found similarly high heterozygosity ( $>0.7$ ) in TLR1, TLR2 and TLR5, and significant results of Fu and Li's D\* test for TLR2 and TLR5, and positive values

of Tajima's D in LRRs in TLR1 and TLR2. Together with significant association between certain variants of TLR1 and TLR5 and susceptibility to pathogens, our results indicate balancing selection acting on these loci.

Differences in pathogens community between populations may explain why it is difficult to detect balancing selection in a macro evolutionary scale. On a large scale the populations are likely to be affected by different pathogen species, whereas on a small scale they are likely to share similar species differing in subtler ways, for instance in their antigenicity. Indeed, compared to directional selection reflecting large-scale differences, the effect of balancing selection is often subtle<sup>16</sup> and seldom persists over a long time<sup>17</sup>. However, as indicated in our study, it seems to be an important factor driving evolution of MHC genes in a micro evolutionary scale.

Directional selection observed within TLRs has been often ascribed to the conserved structure of the PAMPs<sup>54</sup> but these pathogen-derived motifs are not invariable. For instance, TLR4 recognizes lipid A, a conserved part of lipopolysaccharides (LPS) forming the cell wall of Gram negative bacteria. The outer part of the LPS is highly variable, and its polymorphism differentiates strains into particular serotypes<sup>55</sup>. What is more, some pathogens can modify their lipid A to avoid recognition by TLR4, which gives an opportunity for parasite-driven selection<sup>56</sup>.

TLR5, in which we found signatures of balancing selection, recognizes bacterial flagellin, a molecule that is crucial for pathogen motility. The N- and C-terminal parts of the flagellin are conserved but the central region of the protein, which has antigenic properties, is highly variable both within and among bacteria species<sup>57,58</sup>. We found significant associations between allele TLR5\*02 and susceptibility to infection with *Bartonella sp.*, a flagellate bacteria. Samples collected from rodents living in our study area revealed that *Bartonella* strains are highly recombinant<sup>59</sup>, and some of them displayed high variability in virulence gene VirB5<sup>60</sup>. Such a variability gives an opportunity for a micro evolutionary interplay between bacterial variants and TLR5 alleles, in a similar manner as in the case of MHC genes<sup>39</sup>, which might explain the observed signatures of balancing selection in this locus.

**TLR polymorphism and susceptibility to infections.** Functional links between genetic variants and susceptibility are crucial for determining recent or ongoing selection processes<sup>50</sup>, as the ability of a host to resist infection is expected to have a direct impact on its fitness. Except for association studies in humans and livestock, evidence from free-living species is rare and no previous study has considered several TLR loci. In bank voles, TLR2 haplotypes associated with infection with *Borrelia* were identified<sup>7</sup>, and an association between TLR2 and cestode burden was observed in field voles<sup>43</sup>. Here, we found a significant effect of TLR1 and TLR5 alleles on the risk of infection with *Bartonella sp.*

TLR1 presents bacterial lipoproteins, and together with TLR2 and TLR6 they act as a functional subgroup whose members interact with each other, forming heterodimers<sup>61</sup>. Thus, although none of the TLR2 and TLR6 variants remained significant after correcting for multiple comparisons, we may interpret the associations involving TLR1 as a functional outcome of the polymorphism within the TLR1-2-6 group, which underlies the complexity of the associations between components of the innate immunity.

The allele TLR5\*02 increased the risk of infection with *Bartonella*. The presence of alleles associated with increased susceptibility has been widely discussed in the case of MHC, and several explanations have been proposed. Under balancing selection, pathogens are likely to adapt to the most frequent allele, and the rare variants are expected to be associated with resistance<sup>62</sup>. However, the process of removing "susceptible" alleles from a population is slow<sup>63</sup>. Alternatively, the presence of a "susceptible" allele in a population may result from the fact that individuals without this variant did not survive an infection and thus could not be sampled, while hosts with the allele persisted in a population despite heavy pathogen load<sup>64</sup>.

The pathogen community within a host is relatively stable, and the presence of one species may alter the risk of infection with another<sup>65,66</sup>. For instance, in field voles, the risk of infection with *Babesia* increased when an animal was already infected with *Anaplasma*, but voles infected with *Babesia* were more resistant to *Bartonella*<sup>67</sup>. On a molecular level, these associations involve interactions between different pathogen species, and between pathogens and the host immune system. For instance, nematodes may modulate host immune response so that it affects potentially competing species<sup>66</sup> or larval stages of their own species in a process called concomitant immunity<sup>68</sup>, and negative associations between nematode species infecting free living rodents have been reported<sup>18,66</sup>. Microbial pathogens may evade the immune response by modifying their surface proteins interacting with components of the host immune system or through interference with downstream signalling pathways<sup>55</sup>. Moreover, PAMPs derived from a single pathogen are often recognized by various TLRs<sup>69</sup>, and activation of a given TLR by one pathogen may alter its reactivity to another<sup>70</sup>. Thus, the association between the presence of TLR5\*aa02 and susceptibility to infection with *Bartonella* reported in the current paper might result from interactions of this allele with other components of the immune system and/or other pathogens, which were either not recognized in the current study or were too rare to be included in the models.

Another factor that might influence susceptibility is previous exposure to the pathogen. The population studied in the current paper was sampled when the prevalence of *Bartonella* was the highest<sup>71</sup>. Given the relatively short lifespan of a vole (3–4 months) and seeing as voles usually get infected when several weeks old and the infection lasts 1.5 months on average<sup>71</sup>, we may assume that among the individuals classified as uninfected in the current study, only a few were previously exposed to the pathogen and had had enough time to clear from infection.

In the current paper we found weak evidence for the role of TLRs in immunity against nematodes. The presence of the allele TLR1\*aa10 decreased the risk of infection with the nematode *A. tetraptera* by 5-fold, yet this effect was insignificant after correcting for multiple comparisons. Although TLRs are not capable of recognizing motifs derived from multicellular pathogens, they may contribute to the immune response against nematodes through interaction with their *Wolbachia*-like bacterial endosymbionts<sup>72,73</sup>, and associations between TLR2 variants and susceptibility to cestodes was found in common voles<sup>43</sup>. Moreover, helminth infections limit the density of red blood cell dwelling microparasites through anemia and suppression of an inflammatory Th-1 type immune response<sup>74</sup>. Taken together, although components of the innate immunity do not primarily interact with

multicellular pathogens, the presence of helminth infections may affect evolution of the TLR genes and should be considered in future studies.

Although associations between single-nucleotide changes in TLR genes and diseases were detected in human studies (e.g.<sup>6</sup>), we did not find an effect of any SNP on the risk of infection. In free-living species, genotype at several SNPs within cytokines altered the risk of infection<sup>43</sup>. Contrary to the current study, where we covered 47–99% of the coding sequences, study<sup>43</sup> genotyped only a few SNPs per locus. Thus, these might not reflect the overall polymorphism. In our work, haplotypes usually differed from each other in several positions, which suggests that the effect of a single substitution in a protein composed of several hundreds of amino-acids may be undetectable. Moreover, the significant effects of haplotypes underlie the functional importance of the combination of SNPs within a sequence.

**Selection acting on viral TLRs.** In the present study we were not able to test for associations between TLR7 or TLR9 and susceptibility to infection due to low variability at these loci, thus we were able to analyse the selection pattern acting on these loci only based on the results of neutrality tests.

TLR7 recognizes ssRNA from viruses, whereas TLR9 binds CpG DNA from viruses but also bacteria. We found differences in the polymorphisms of TLR7 and TLR9 compared to the remaining loci involved in the response against bacterial or protozoan pathogens. TLR7 was characterized by a low number of segregating sites, low haplotype diversity, and consistent excess of homozygotes in both studied sites. This characteristic, together with negative Tajima's D and neutrality index  $NI > 1$  (albeit with both values not significant), may suggest purifying selection. TLR9 also had a low number of segregating sites and low haplotype diversity but significant  $NI < 1$  suggested positive selection. In both loci we observed high and significant  $F_{st}$  values, indicating population differentiation ( $F_{st} = 0.11$  in TLR7 and 0.38 in TLR9). However, the results of neutrality tests may be affected by a low overall variability in these loci and high differences in allele frequencies between sites, thus they should be interpreted with caution.

Differences in both the strength and direction of selection acting on TLRs presenting virus-derived nucleic acids compared to TLR loci presenting bacterial PAMPs have previously been reported<sup>8,14</sup>, and the low polymorphism of TLR7 has been attributed to low diversifying selective pressures acting on viral ssRNA<sup>14</sup>. In an alternative explanation<sup>8</sup>, prevention of self-reactivity has been proposed as the main reason for low polymorphism within TLRs interacting with viral PAMPs: these TLRs are expressed in the endoplasmic reticulum, so there is a risk of an immune response against self nucleic acids.

TLR3, 7, and 9 are not the only pattern recognition receptors that interact with viruses. Viruses that have already entered into the cytoplasm may be detected by RIG-I-like receptors (RLRs), such as RIG-I and MDA5, which contain a helicase domain capable of binding to viral RNA. Their activation induces pro-inflammatory innate immune response, and each RLR recognize different viral patterns<sup>75</sup>. Several viral proteins that inhibit RLRs signaling have been identified suggesting that pathogens evolve to avoid recognition by RLR-induced pathway<sup>75</sup>. Although there is no studies comparing selective pressure acting on both classes of these viral-sensing PRRs, we may hypothesise that strong selective pressure on RIG-I and MDA5 may weaken selection acting on TLR7 or TLR9. Moreover, nucleic acids are not the only viral patterns recognized by components of the innate immunity. For instance, highly variable proteins expressed on a viral envelope are recognized by TLR2 and TLR4<sup>69</sup>, what may also contribute to the relaxed selection acting on TLR7 and TLR9.

On the other hand, low polymorphism of TLR9 may contribute to the biochemical properties of the molecule, which undergoes considerable conformational change after binding a ligand<sup>76</sup>. To preserve this ability, a large part of the sequence has to be conserved. Hence, the polymorphism in TLR9 may be explained by mechanisms maintaining the stereochemical properties of the molecule<sup>61</sup>, interactions with other components of the immune system, or preventing self-reactivity, rather than selective pressure from pathogens. Testing this hypothesis will require further studies using laboratory animals, including infections with RNA and DNA viruses.

## References

1. Acevedo-Whitehouse, K. & Cunningham, A. A. Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology and Evolution* **21**, 433–438 (2006).
2. Jepson, A. *et al.* Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infections and Immunity* **65**, 872–876 (1997).
3. Behnke, J. M. *et al.* Chasing the genes that control resistance to gastrointestinal nematodes. *Journal of Helminthology* **77**, 99–109 (2003).
4. Gavan, M. K., Oliver, M., Douglas, A. & Pieltney, S. B. Gene dynamics of toll-like receptor 4 through a population bottleneck in an insular population of water voles (*Arvicola amphibius*). *Conservation Genetics* **16**, 1181–1193 (2015).
5. Wenzel, M. A. & Pieltney, S. B. Digging for gold nuggets: uncovering novel candidate genes for variation in gastrointestinal nematode burden in a wild bird species. *Journal of Evolutionary Biology* **28**, 807–825 (2015).
6. Skevaki, C., Pararas, M., Kostelidou, T., Tsakris, A. & Ooutsias, J. G. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious diseases. *Clinical and Experimental Immunology* **180**, 165–77 (2015).
7. Tschirren, B. *et al.* Polymorphisms at the innate immune receptor TLR2 are associated with *Borrelia* infection in a wild rodent population. *Proceeding of the Royal Society B* **280**, 20130364 (2013).
8. Wlasiuk, G. & Nachman, N. W. Adaptation and constraint at Toll-like receptors in primates. *Molecular Biology and Evolution* **27**, 2172–86 (2010).
9. Nakajima, T. *et al.* Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* **60**, 727–35 (2008).
10. Areal, H., Abrantes, J. & Esteves, P. J. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evolutionary Biology* **11**, 368 (2011).
11. Alcaide, M. & Edwards, S. V. Molecular evolution of the Toll-like receptor multigene family in birds. *Molecular Biology and Evolution* **28**, 1703–15 (2011).
12. Babik, W. *et al.* Constraint and Adaptation in newt Toll-Like Receptor Genes. *Genome Biology and Evolution* **7**, 81–95 (2015).
13. Tschirren, B., Råberg, L. & Westerdahl, H. Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. *Journal of Evolutionary Biology* **24**, 1232–40 (2011).



14. Fornusková, A. *et al.* Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). *BMC Evolutionary Biology* **13**, 194 (2013).
15. Ferrer-Admetlla, A. *et al.* Balancing Selection Is the Main Force Shaping the Evolution of Innate Immunity Genes. *Journal of Immunology* **181**, 1315–1322 (2008).
16. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annual Review in Genetics* **47**, 97–120 (2013).
17. Charlesworth, D. Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. *PLoS Genetics* **2**, e64 (2006).
18. Bajer, A. *et al.* Medium-term temporal stability of the helminth component community structure in bank voles (*Clethrionomys glareolus*) from the Mazury Lake District region of Poland. *Parasitology* **130**, 213–228 (2005).
19. Kloch, A., Babik, W., Bajer, A., Siński, E. & Radwan, J. Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Molecular Ecology* **19**(Suppl 1), 255–265 (2010).
20. Heinriksen, S. & Pohlenz, J. Staining of cryptosporidia by modified Ziehl-Nielsen technique. *Acta Veterinaria Scandinavica* **22**, 594–596 (1981).
21. Norman, A. F., Regnery, R., Jameson, P., Greene, C. & Krause, D. C. Differentiation of Bartonella-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. *Journal of Clinical Microbiology* **33**, 1797–1803 (1995).
22. Herwaldt, B. L. *et al.* Molecular characterization of a non-Babesia divergens organism causing zoonotic babesiosis in Europe. *Emerging Infectious Diseases* **9**, 942–948 (2003).
23. Persing, D. H. *et al.* Detection of Babesia microti by polymerase chain reaction. *Journal of Clinical Microbiology* **30**, 2097–2103 (1992).
24. Zieliński, P., Stuglik, M. T., Dudek, K., Konczal, M. & Babik, W. Development, validation and high-throughput analysis of sequence markers in nonmodel species. *Molecular Ecology Resources* **14**, 352–360 (2014).
25. Wu, T. D. & Watanabe, C. K. GMAP, a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
26. Quinlan, A. R. & Hall, I. M. BEDTools, a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–841 (2010).
27. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler Transformation. *Bioinformatics* **26**, 589–95 (2010).
28. Garrison, E., Marth, G. Haplotype-based variant detection from short-read sequencing. Unpublished work, arXiv:1207.3907 (2012).
29. Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**, 443–451 (2011).
30. Stephens, M., Smith, N. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–989 (2001).
31. Offord, V., Coffey, T. J. & Werling, D. A web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. *Developmental and Comparative Immunology* **34**, 1035–41 (2010).
32. Librado, P. & Rozas, J. DNASPv5, A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
33. Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W. & Prodöhl, P. A. diveRsity, An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**, 782–788 (2013).
34. Excoffier, L., Laval, G. & Schneider, S. Arlequin ver., 3.0, An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**, 47–50 (2005).
35. Rousset, F. Genepop'007, a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106 (2008).
36. Tajima, F. The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601 (1989).
37. Fu, Y.-X. & Li, W.-H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
38. McDonald, J. H. Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Molecular Biology and Evolution* **15**, 377–384 (1998).
39. Ewens, W. J. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112 (1972).
40. Watterson, G. A. The homozygosity test of neutrality. *Genetics* **88**, 405–417 (1978).
41. Piertney, S. B. & Oliver, M. K. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21 (2006).
42. R Core Team R, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/> (2016).
43. Turner, A. K., Begon, M., Jackson, J. A., Bradley, J. E. & Paterson, S. Genetic Diversity in Cytokines Associated with Immune Variation and Resistance to Multiple Pathogens in a Natural Rodent Population. *PLoS Genetics* **7**, e1002343 (2011).
44. Kloch, A., Baran, K., Buczek, M., Konarzewski, J. & Radwan, J. MHC influences parasite load and winter survival in the root vole *Microtus oeconomus*. *Evolutionary Ecology* **27**, 635–653 (2013).
45. Gonçalves da Silva, A. & Russello, M. A. IRel, software for implementing pairwise relatedness estimators and evaluating their performance. *Conservation Genetics Resources* **3**, 69–71 (2011).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. **57**, 289–300 (1995).
47. Wenzel, M. A., James, M. C., Douglas, A. & Piertney, S. B. Genome-wide association and genome partitioning reveal novel genomic regions underlying variation in gastrointestinal nematode burden in a wild bird. *Molecular Ecology* **24**, 4175–92 (2015).
48. Wang, Y. *et al.* RSEQ, a statistical method for quantification of isoform-specific transcriptome using transcriptome-wide structural profiling data. Unpublished work, [www.biorxiv.org/content/early/2016/06/18/043232](http://www.biorxiv.org/content/early/2016/06/18/043232) (2016).
49. Chang, C. C. *et al.* Second-generation PLINK, rising to the challenge of larger and richer datasets. *GigaScience*, **4** (2015).
50. Turner, A. K., Begon, M., Jackson, J. A. & Paterson, S. Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology* **21**, 1632–46 (2012).
51. Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H. & Råberg, L. Contrasting patterns of diversity and population differentiation at the innate immunity gene toll-like receptor 2 (TLR2) in two sympatric rodent species. *Evolution* **66**, 720–31 (2012).
52. Barreiro, L. B. *et al.* Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genetics* **5**, e1000562 (2009).
53. Mukherjee, S., Sarkar-Roy, N., Wagener, D. K. & Majumder, P. P. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proceedings of the National Academy of Sciences USA* **106**, 7073–7078 (2009).
54. Janssens, S. & Beyaert, R. Role of Toll-like receptors in pathogen recognition. *Clinical Microbiology Reviews* **16**, 637–646 (2003).
55. Finlay, B. B. & McFadden, G. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. *Cell* **124**, 767–782 (2006).
56. Portnoy, D. A. Manipulation of innate immunity by bacterial pathogens. *Current Opinion in Immunology* **17**, 25–28 (2005).
57. Macnab, R. M. Flagella and motility. In: Neidhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umberger, H. E. (Eds). *Escherichia coli and Salmonella: cellular and molecular biology*. Washington. ASM Press, pp. 123–145 (1996).
58. Winstanley, C. & Morgan, J. A. W. The bacterial flagellin gene as a biomarker for detection, population genetics and epidemiological analysis. *Microbiology* **143**, 3071–3084 (1997).

59. Paziewska, A., Harris, P. D., Zwolińska, L., Bajer, A. & Siński, E. Recombination within and between species of the alpha proteobacterium *Bartonella* infecting rodents. *Microbial Ecology* **61**, 134–45 (2011).
60. Paziewska, A. Diversity of blood parasites of genus *Bartonella* in wild rodents in Mazury Lakes District. in Polish. *Wiadomości Parazytologiczne* **57**, 53–4 (2011).
61. Botos, I., Segal, D. M. & Davies, D. R. The Structural Biology of Toll-Like Receptors. *Structure* **19**, 447–459 (2011).
62. Spurgin, L. G. & Richardson, D. S. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of Royal Society B* **277**, 979–988 (2010).
63. Apanius, V., Penn, D., Slev, P. R. & Potts, W. K. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17**, 179–224 (1997).
64. Biedrzycka, A. & Kloch, A. Development of novel associations between MHC alleles and susceptibility to parasitic infections in an isolated population of an endangered mammal. *Infections, Genetics and Evolution* **44**, 210–217 (2016).
65. Knowles, S. C. L. *et al.* Stability of within-host-parasite communities in a wild mammal system. *Proceedings of the Royal Society B* **280**, 20130598 (2013).
66. Lello, L., Boag, B., Fenton, A., Stevenson, I. R. & Hudson, P. J. Competition and mutualism among gut helminths of a mammalian host. *Nature* **428**, 840–844 (2004).
67. Telfer, S. *et al.* Species Interactions in a Parasite Community Drive Infection Risk in a Wildlife Population. *Science* **330**, 243–246 (2010).
68. Brown, P. S. & Grenfell, B. T. An unlikely partnership: parasites, concomitant immunity and host defence. *Proceedings of Royal Society B* **268**, 2543–2549 (2001).
69. Uematsu, S. & Akira, S. Toll-like receptors (TLRs) and their ligands. *Handbook of Experimental Pharmacology* **183**, 1–20 (2008).
70. Trinchieri, G. & Sher, A. Cooperation of Toll-like receptor signals in innate immune defence. *Nature Reviews Immunology* **7**, 179–190 (2007).
71. Paziewska, A., Harris, P. D., Zwolińska, L., Bajer, A. & Siński, E. Differences in the ecology of *Bartonella* infections of *Apodemus flavicollis* and *Myodes glareolus* in a boreal forest. *Parasitology* **139**, 881–93 (2012).
72. Taylor, M. J., Cross, H. F. & Bilo, K. Inflammatory responses induced by the filarial nematode *Brugia malayi* are mediated by lipopolysaccharide-like activity from endosymbiotic *Wolbachia* bacteria. *Journal of Experimental Medicine* **191**, 1429–36 (2000).
73. Punkosdy, G. A., Addiss, D. G. & Lammie, P. J. Characterization of antibody responses to *Wolbachia* surface protein in humans with lymphatic filariasis. *Infections and Immunity* **71**, 5104–14 (2003).
74. Graham, A. Ecological rules governing helminth-microparasite coinfection. *Proceedings of the National Academy of Sciences USA* **105**, 566–570 (2008).
75. Yoneyama, M. & Fujita, T. Function of RIG-I-like receptors in antiviral innate immunity. *The Journal of Biological Chemistry* **282**, 15315–15318 (2007).
76. Latz, E. *et al.* Ligand-induced conformational changes allosterically activate Toll-like receptor 9. *Nature Immunology* **8**, 772–779 (2007).

## Acknowledgements

The work was supported by grant no. DEC-2012/07/B/NZ8/00058 from the Polish National Science Centre to A.K. We are thankful to Dr. hab W. Babik who provided access to an Illumina MiSeq platform, and to K. Dudek who prepared the Nextera library. Special thanks to A. Biedrzycka for her valuable comments on the final version of the manuscript. We also would like to thank two anonymous reviewers for their valuable comments that helped to improve the manuscript.

## Author Contributions

A.K. conceived, designed and performed the study, analysed data, and wrote the paper with contributions from D.R.L., M.A.W. and S.B.P. D.R.L. analysed genetic data; O.M. contributed to the data analysis and wrote scripts; R.W.F. analysed infection with blood pathogens.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-26672-2>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018